

Evaluation von Disease-Management-Programmen – Aktuelle Defizite, Anforderungen, Methoden

Ferdinand M. Gerlach¹, Martin Beyer¹, Joachim Szecsenyi² und Heiner Raspe³

¹Institut für Allgemeinmedizin, Christian-Albrechts-Universität zu Kiel

²Sektion Allgemeinmedizin und Versorgungsforschung, Universität Heidelberg

³Institut für Sozialmedizin, Medizinische Universität Lübeck

ZUSAMMENFASSUNG

Die Evaluation von Disease-Management-Programmen ist notwendig und hierzulande im Sozialgesetzbuch V obligatorisch vorgeschrieben. Bisher besteht jedoch noch weitgehende Unklarheit über die konkrete Ausgestaltung geeigneter Strategien zur Programmevaluation. Anhand der wissenschaftlichen Begleitung der Diabetesvereinbarungen in Nordrhein werden zunächst exemplarisch aktuelle Defizite in der Evaluation bereits bestehender Programme für chronisch Kranke dargestellt. Auf der Basis internationaler Erfahrungen werden dann sinnvolle Anforderungen diskutiert, die bei der Evaluation von Wirksamkeit und Nutzen der Programme erfüllt sein sollten. Als Ausweg aus dem methodischen Dilemma unterschiedlich geeigneter Studiendesigns wird nach dem Vorbild des britischen Medical Research Council ein vierstufiges Modell vorgeschlagen, das eine angemessene Evaluation auch komplexer Interventionen erlaubt.

Sachwörter: Disease Management, Programmevaluation, Qualitätssicherung, Evaluationsmethoden

Die Rechtsverordnung, in der die Anforderungen an DMPs für Patienten mit Diabetes mellitus Typ 2 festgelegt sind, lässt allgemeine Grundlagen und Ziele einer Evaluation aus Sicht des Verordnungsgebers erkennen (Übersicht 1).

Durch Vorgabe und Überprüfung methodischer Kriterien soll das Bundesversicherungsamt darauf hinwirken, dass die Evaluationen unterschiedlicher Programme diagnosebezogen vergleichbar sind.^c

EINE EVALUATION IST SINNVOLL UND OBLIGATORISCH VORGESCHRIEBEN

Nach dem Willen des Gesetzgebers (§ 137 f SGB V) werden derzeit in Deutschland flächendeckend Disease-Management-Programme (DMP) für chronisch Kranke etabliert. Die Anforderungen an strukturierte Programme für Patienten mit Diabetes mellitus Typ 2, Brustkrebs und Koronarer Herzkrankheit sind bereits in entsprechenden Rechtsverordnungen des Bundesministeriums für Gesundheit und Soziale Sicherung (BMGS) zur Änderung der Risikostruktur-Ausgleichsverordnung festgelegt.^a In Vorbereitung sind Programme für

Diabetes mellitus Typ 1, Asthma und COPD. Die Erwartungen an die Effekte dieser mit enormem Aufwand verbreiteten Programme sind hoch. Um deren Wirksamkeit und Nutzen, insbesondere ihre tatsächlichen Auswirkungen auf die Versorgungsqualität beurteilen zu können, ist eine Evaluation unverzichtbar. In § 137f Abs. 2 Nr. 6 ist denn auch eine obligatorische Evaluation der Programme vorgesehen. Die dabei entstehenden Kosten müssen die Krankenkassen übernehmen.^b

^b„Die Krankenkassen oder ihre Verbände haben eine externe Evaluation der Programme nach Absatz 1 durch einen vom Bundesversicherungsamt im Benehmen mit der Krankenkasse oder dem Verband auf deren Kosten bestellten unabhängigen Sachverständigen auf der Grundlage allgemein anerkannter wissenschaftlicher Standards zu veranlassen, die zu veröffentlichen ist.“ (§ 137f Abs. 4 SGB V)

DIE REALITÄT – STAND DER EVALUATION VON PROGRAMMEN FÜR CHRONISCH KRANKE AM BEISPIEL DER DIABETESVEREINBARUNGEN IN NORDRHEIN

Anhand des Berichts über die wissenschaftliche Begleitung der Diabetesvereinbarungen in Nordrhein (2) werden im folgenden aktuelle Defizite der Evaluation bereits bestehender Programme für chronisch Kranke exemplarisch dargestellt. Die vorliegende Publikation ist als

^c„Das Bundesversicherungsamt hat durch Vorgaben methodischer Kriterien darauf hinzuwirken, dass die Evaluationen unterschiedlicher Programme diagnosebezogen vergleichbar sind.“ (§ 28 g der 4. Verordnung zur Änderung der Risikostruktur-Ausgleichsverordnung vom 27. Juni 2002)

^a4. RSA-ÄndV vom 27.06.2002, zuletzt geändert am 27.12.2002 und 7. RSA-ÄndV vom 21.04.2003



Übersicht 1: Ziele und Grundlagen der Evaluation nach § 137f. Abs. 2 Nr. 6 SGB V (laut Vierter Verordnung zur Änderung der Risikostruktur-Ausgleichsverordnung vom 27. Juni 2002, Anlage 1: Anforderungen an strukturierte Behandlungsprogramme für Diabetes mellitus Typ 2)

„Grundziele der Evaluation sind die Überprüfung

- der Erreichung der Ziele des strukturierten Behandlungsprogramms,
- der Einhaltung der Einschreibekriterien,
- der Sicherstellung einer qualitativ hochwertigen Versorgung nicht am Programm beteiligter Patienten mit programmgleichen oder anderen Diagnosen sowie
- der ökonomischen Effizienz.

Die Ziele des Programms ergeben sich aus den Anforderungen gemäß § 137f. Abs. 2 Satz 2 Nr. 1 des Fünften Buches Sozialgesetzbuch (Anforderungen an die Behandlung nach evidenz-basierten Leitlinien unter Berücksichtigung des jeweiligen Versorgungssektors) und § 137f. Abs. 2 Satz 2 Nr. 2 des Fünften Buches Sozialgesetzbuch (durchzuführende Qualitätssicherungsmaßnahmen) sowie den Vereinbarungen zu den Qualitätssicherungsmaßnahmen.

Grundlage der Evaluation bilden die für den Evaluationszeitraum relevanten versichertenbezogenen Dokumentationen nach § 28f. der Risikostruktur-Ausgleichsverordnung, alle Leistungsdaten sowie Abrechnungsdaten der teilnehmenden Leistungserbringer für die im Evaluationszeitraum eingeschriebenen Versicherten. Die Daten werden für die Zwecke der Evaluation pseudonymisiert.

Bei der Bewertung der Wirksamkeit des strukturierten Behandlungsprogramms ist zwischen der Funktionsfähigkeit des Programms und seiner Auswirkung auf die Versorgungslage zu unterscheiden:

- Bei der Beurteilung der Funktionsfähigkeit des Programms sind insbesondere die Anforderungen gemäß § 137f. Abs. 2 Satz 2 Nr. 1 und 2 des Fünften Buches Sozialgesetzbuch einschließlich des Verfahrens der Vereinbarung individueller Therapieziele zu evaluieren.
- Gradmesser für die Auswirkung auf die Versorgungslage ist die Veränderung der Ausprägungen von Parametern der Prozess- und Ergebnisqualität des Mindest-Datensatzes relativ zu den ermittelten Ausgangswerten. Die Möglichkeiten des Vergleiches zu einer Kontrollgruppe nicht eingeschriebener Versicherter/nicht teilnehmender Leistungserbringer sind zu prüfen.

Die Evaluation kann auf der Grundlage einer repräsentativen Stichprobe der eingeschriebenen Versicherten erfolgen; sie ermöglicht eine versichertenbezogene Verlaufsbetrachtung über den Evaluationszeitraum.

Der Prozentsatz sowie die Versichertenstruktur der teilnehmenden Versicherten je Krankenkasse ist zu berücksichtigen. Versicherte, die das strukturierte Behandlungsprogramm freiwillig oder durch Ausschluss verlassen, sind besonders zu würdigen.

Die Evaluation soll auch subjektive Ergebnisqualitätsparameter (Lebensqualität, Zufriedenheit) auf der Basis einer einmaligen Stichproben-Befragung bei eingeschriebenen Versicherten mindestens jeweils zu Beginn und zum Ende des Evaluationszeitraums umfassen. Hierfür ist ein Adressmitteilungsverfahren durch die Krankenkasse vorzusehen.

Unter der Berücksichtigung der benötigten Datenbasis können die Vertragspartner vereinbaren, inwieweit zu evaluieren ist, ob die Programme Auswirkungen auf die Versorgung von nicht eingeschriebenen Versicherten haben.

Die Evaluation beginnt ein halbes Jahr nach Akkreditierung des Programms und umfasst einen Zeitraum von drei Jahren. Jährliche Zwischenberichte sind an die Krankenkassen zu liefern und von diesen binnen acht Wochen zu veröffentlichen.“

solche unbedingt zu begrüßen, da sie hierzulande zu den ersten gehört, in denen die Evaluation eines strukturierten Programms zur Versorgung chronisch Kranker systematisch aufbereitet und öffentlich zugänglich gemacht wurde.

Einer breiteren Fachöffentlichkeit gegenüber wurde die vom Zentralinstitut für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland durchgeführte Evaluation als die „weltweit größte wissenschaftliche Evaluation einer strukturierten Diabetiker-Versorgung“^d vorgestellt. Die Autoren selbst weisen in einer Vorbemerkung darauf hin, dass „die mit den Diabetesvereinba-

rierten Diabetiker-Versorgung“^d vorgestellt. Die Autoren selbst weisen in einer Vorbemerkung darauf hin, dass „die mit den Diabetesvereinba-

^dDeutsches Ärzteblatt 99: A2826 vom 25.10.2002



rungen Nordrhein gemachten Erfahrungen hinsichtlich Akzeptanz und der in kurzer Zeit erreichten Ergebnisse einen Maßstab bilden, an denen sich die Disease Management Programme alsbald messen lassen müssen.“ (2, S. 12).

In der Tat ist es lohnend, sich mit dem hier erreichten Stand auseinanderzusetzen. Bei genauerer Betrachtung fallen nämlich einige methodische Probleme auf, die für die zu überwindenden Schwierigkeiten einer Evaluation von DMPs exemplarisch sein könnten:

Studienmethodik

Es handelt sich um eine unkontrollierte, nicht randomisierte Beobachtungsstudie, in der Längs- und Querschnittsauswertungen zur Evaluation der Interventionseffekte kombiniert werden (Level of Evidence: III). Eine Kausalität zwischen gemessenen Effekten und den Auswirkungen des Programms lässt sich infolge des Fehlens einer Kontrollgruppe prinzipiell nicht belegen.

Ziele und Fragestellungen

Ziele und Fragestellungen der Evaluationsstudie (S. 44^e) stimmen nur teilweise mit den Zielen der Diabetesvereinbarungen (S. 15) überein. Auffällig ist vor allem, dass drei von vier „wesentlichen Zielen“ der Vereinbarungen überhaupt nicht untersucht wurden: Veränderungen der Lebensführung der Patienten^f, Kosten(einsparungen), Vermeidung stationärer Behandlung.

Beobachtungsdauer

Der Beobachtungszeitraum (3 Jahre bzw. 12 Quartale) lässt grundsätzlich nur begrenzte Aussagen zum Verlauf einer chronischen Erkrankung wie Diabetes mellitus zu. Insbesondere evtl. Verbesserungen an-

hand harter Endpunkte (terminale Niereninsuffizienz, kardiovaskuläre Folgeerkrankungen etc.) können generell (noch) nicht erwartet werden. Neben diesem prinzipiellen Problem kommt hinzu, dass viele Praxen (und deren Patienten) im Beobachtungszeitraum (Quartale I/99 bis IV/01) erst sukzessive rekrutiert bzw. eingeschlossen wurden. Aus 2.207 Zugängen und 300 Abgängen resultierten zwangsläufig stark unterschiedliche Beobachtungszeiträume und Messhäufigkeiten (S. 48).

Selektionseffekte

Die Teilnahmequote hausärztlicher Praxen erreichte mit 45% einen durchaus erfreulichen Wert. Trotzdem wird deutlich, dass die teilnehmenden Praxen eine Selektion von Praxen mit „deutlich mehr Diabetikern“ darstellen (S. 46). Unklar bleibt, wie viele der insgesamt betroffenen Diabetiker einer Praxis tatsächlich erreicht wurden und wie sich die Schweregrade ihrer Erkrankung verteilen. Aufgrund der unklaren Struktur der Patientenkollektive fällt insbesondere die Einordnung der im Bericht gegenübergestellten Befunde von Patienten in hausärztlichen Praxen und Schwerpunktpraxen schwer.

Messmethoden

Naturgemäß handelt es sich bei der vorgelegten Studie um eine Analyse von Sekundärdaten, zumeist um von Praxen selbst gemessene (RR) bzw. dokumentierte Werte aus unterschiedlichen Laboren. Dabei fällt auf, dass zur Bestimmung des HbA1c in unterschiedlichen Laboren offensichtlich unterschiedliche Bestimmungsmethoden (S. 58) verwendet wurden, was zwangsläufig zu nicht miteinander vergleichbaren Normwerten führt. Noch stärker wäre die Vergleichbarkeit eingeschränkt, wenn – wie offenbar im Rahmen der Vergütungsregelungen vorgesehen – zum Teil tatsächlich auch HbA1(!)-Werte dokumentiert wurden (S. 20). Eine erhebliche Einschränkung der Beurteilbarkeit er-

gibt sich auch aus der Tatsache, dass Hypertoniker nicht gesondert betrachtet wurden. Obwohl gerade der Einstellung des Blutdrucks bei Diabetikern eine zentrale Bedeutung zukommt, lassen sich hier eventuelle Therapieeffekte bzw. die Einstellungsqualität leider nicht differenzieren.

Neben diesen methodischen Problemen, die auch bei der zukünftigen Evaluation von DMPs zu berücksichtigen sind und nach Möglichkeit überwunden werden sollten, enthält der Bericht aus Nordrhein auch noch einige auffällige Ergebnisse, die auf mögliche Fallstricke hinweisen. So sind z.B. nur 15,5% der Patienten, die laut bestehendem Vertrag eine „Schnittstelle“ zur Überweisung erreicht hatten, auch tatsächlich an eine Schwerpunktpraxis überwiesen worden (S. 80). Hier sind u.a. folgende Fragen offen: Werden hier die „richtigen“ Patienten überwiesen? Verhindern Hausärzte mit Augenmaß evtl. eine ansonsten vertraglich induzierte Überversorgung?

Es fällt weiter auf, dass nur ein kleines Kollektiv (3,5% aller erfassten Patienten), das anfänglich einen deutlich erhöhten HbA1c (Messwert über 10%) aufwies, eine deutliche Senkung zeigte (S. 116). Ebenfalls nur ein kleines Kollektiv der Diabetiker (4,8% aller erfassten Patienten), das anfänglich einen erhöhten systolischen Blutdruck (über 160 mmHg) aufwies, zeigte eine deutliche Senkung (S. 115). In diesem Zusammenhang ist erkennbar, dass die Einstellung (HbA1c) der eingeschlossenen Patienten ohnehin zumeist schon bei Aufnahme in das Programm außergewöhnlich gut war. Offene Fragen hier: Handelt es sich dabei um ein Zeichen einer ohnehin (auch ohne Vertrag) guten Stoffwechseleinstellung oder um den Ausdruck eines Selektionseffektes besonders motivierter Praxen bzw. Patienten (S. 46/47), oder ist evtl. eine zu weite Indikationsstellung infolge der im Vertrag gesetzten Vergütungsanreize ausschlaggebend (S. 46)?

^eDiese und alle folgenden Seitenangaben beziehen sich auf den Bericht von Altenhofen et al. 2002 (2)

^fEine Erhebung der Patientensicht (subjektive Behandlungszufriedenheit) ist nach Angaben der Autoren (S. 45) kurzfristig geplant.



Fazit: Der vorgestellte Evaluationsansatz ist ein erster Schritt in die richtige Richtung. Inwieweit die Diabetesvereinbarungen in Nordrhein ihre selbstgesteckten Ziele erreicht haben, bleibt aber infolge methodischer Schwächen bisher noch unklar. Wie die Frage nach Wirksamkeit und Nutzen von Programmen für chronisch Kranke zukünftig beantwortet werden könnte, soll im folgenden gezeigt werden.

DER ANSPRUCH – ANFORDERUNGEN AN EINE SINNVOLLE EVALUATION

Ausgehend von dieser exemplarischen Darstellung der Evaluationsrealität im deutschen Gesundheitswesen sollen zunächst grundlegende Anforderungen an eine sinnvolle Evaluation diskutiert werden. Ein generelles Problem besteht hierbei darin, dass auf der einen Seite eine abschließende summative Evaluation (Wirkungsanalyse) mit einem strengen Studiendesign und auch einer hinreichenden ‚Programmkonstanz‘ zu fordern ist. Auf der anderen Seite soll eine begleitende formative Evaluation (also Prozessevaluation[§]) den Beteiligten bereits im Verlauf der Implementierung die Chance eröffnen, auf Fehlentwicklungen zeitnah reagieren zu können.

DMPs als komplexe Interventionen

DMPs bestehen aus einem komplexen Paket verschiedener Interventionen. Alle DMPs enthalten mindestens folgende verschiedenartigen Elemente: einen Versorgungspfad, verschiedene Schnittstellendefinitionen, Versorgungsaufträge, ein Konzept zur Patienteninformation, ein Dokumentationsraster und eine Vorgabe für Outcome- bzw. Evaluationsdaten. Alle diese Elemente ha-

[§]vgl. zu den Begriffen der summativen und formativen Evaluation weiter unten den Abschnitt ‚Theorie der Evaluation‘

ben unterschiedliche Effektivitätsvermutungen, einen unterschiedlichen Grad der Evidenz und differente (zum Teil synergistische) Auswirkungen. Es wäre verlockend, sich ausschließlich auf harte ‚Outcomes‘ zu konzentrieren, und dabei ein methodisch sehr strenges Design mit definierter Studienstrecke, fixen Endpunkten und unbeeinflusster Kontrollgruppe zu fordern. – Dieses Design würde aber der Komplexität von DMPs kaum gerecht, und die Ergebnisse wären in jedem Fall stark von der konkreten Art und Weise der Implementierung sowie der Wirksamkeit der Einzelmaßnahmen abhängig.

DMPs unterliegen einem Implementierungsprozess und wurden konzipiert, um typische Schwächen der Versorgung chronisch Erkrankter (13, Ziffer 676) zu beheben. *Die Defizite können jedoch nicht ad hoc durch die Vorgabe eines Versorgungspfads, sondern nur sukzessive, bei gleichzeitiger Veränderung von Anreizsystemen überwunden werden.* DMPs, die im Zeitraum von drei Jahren ‚ihr Geld zurückspielen‘ sollen, sind eine Reissbrett-Utopie von Gesundheitsplanern, die insbesondere die Beeinflussbarkeit chronischer Erkrankungen (aber auch die vielerorts bereits erreichte Prozessqualität) unzureichend berücksichtigen. Surrogatparameter (9), die teilweise auch in der zuvor besprochenen Evaluation der Diabetesvereinbarung Nordrhein eine Rolle spielten, sind naturgemäß schneller zu normalisieren als die langfristigen Konsequenzen (z.B. kardiovaskuläre Ereignisse) einer defizienten Versorgung. Unabhängig von der Evidenzbasierung bzw. ‚efficacy‘ der Einzelinterventionen werden letztlich immer verschiedene Akteure mit unterschiedlicher Intensität, Reichweite und Marktdominanz versuchen, diese umzusetzen.

DMPs werden zwangsläufig konfliktierende Perspektiven von Leistungserbringern, Patienten und Kostenträgern berühren. Daher müssen

in den Kriterien und Indikatoren der Evaluation die wesentlichen Versorgungsziele des Programms und die erwünschten Outcomes abgebildet werden.

Evidenzbasierung medizinischer Versorgung

Nach den Kriterien der evidenzbasierten Medizin werden auch an Wirksamkeitsstudien über Versorgungsprogramme und Public-Health-Interventionen strenge Forderungen gestellt (12). Auch hier steht die randomisierte kontrollierte Studie an der Spitze der Evidenzhierarchie, womit nach Möglichkeit eine einheitliche Studienstrecke, klar definierte Intervention, objektive Zielkriterien, weitgehende Ausschaltung von Störeinflüssen sowie natürlich eine Kontrollgruppe gefordert sind. In der internationalen Versorgungsforschung, z.B. bei Krebsfrüherkennungsprogrammen sowie in der Rehabilitation, ist bereits gezeigt worden, dass derartige Studien erfolgreich durchführbar sind. Insbesondere zum Disease Management gibt es internationale Studien und systematische Reviews, die deren Wirksamkeit belegen (u.a. 3, 5, 14) und an die eine Evaluation in Deutschland anknüpfen könnte. Erheblich geringer ist allerdings die Zahl von Studien, die auch gesundheitsökonomische Effekte belegen können. Über die Evaluation von Modellvorhaben in Deutschland selbst sind so gut wie keine wissenschaftlich belastbaren Ergebnisse bekannt geworden – im scharfen Gegensatz zur Gesetzesnorm (§ 65 Satz 2 SGB V): „Der von unabhängigen Sachverständigen zu erstellende Bericht über die Ergebnisse der Auswertung ist zu veröffentlichen“.

Eine methodisch anspruchsvolle summative Evaluation bietet die bestmögliche Voraussetzung für eine unverzerrte und nach klaren Endpunkten bewertbare Beobachtung. Von praktischen Schwierigkeiten der Anwendung auf flächendeckend



eingeführte DMP-Programme einmal abgesehen, muss jedoch gefragt werden, inwieweit ein kontrolliertes Studiendesign zum Zweck der Programmevaluation brauchbar ist, bzw. ob nicht die Verengung der Evaluationsperspektive sowie die ‚Abstinenz‘ von einer korrigierenden Beeinflussung während des Implementierungsprozesses ausschlaggebende Nachteile eines solchen Vorgehens darstellen müssen. Realistischer scheint die Initiierung verschiedener paralleler Kohortenstudien. Mit ihnen ließen sich möglicherweise auch unterschiedlich gestaltete DMPs vergleichen. Fraglich ist, ob sich hierzulande überhaupt Kontroll- oder Vergleichsgruppen von Versicherten, Patienten und Leistungserbringern bilden lassen, die sich gegen DMPs entschieden haben oder von ihnen nicht erreicht wurden. In diesen Fällen wird es schwierig sein, Stichproben zu parallelisieren und sog. Confounder zu kontrollieren.

Auch wenn über das grundsätzliche Studiendesign Einigkeit erzielt werden kann, bleibt die weitere Frage zu klären, wie patientennahe Effekte komplexer Interventionen angemessen abgebildet werden können. Neben ‚harten klinischen Endpunkten‘ und Prozessindikatoren ist zusätzlich ein breites Spektrum von Krankheitsfolgen zu berücksichtigen, wie es z. B. in der International Classification of Functioning, Disability and Health (ICF) der WHO (15) abgebildet wird. Neben konsekutiven Störungen weiterer körperlicher, vor allem aber seelischer Funktionen sind auch Aktivitätseinschränkungen und Störungen der Teilhabe sowie Einschränkungen der gesundheitsbezogenen Lebensqualität einzubeziehen.

Theorie der Programmevaluation

Qualitätskriterien an eine Programmevaluation werden z. B. von der American Evaluation Association in definierten Standards formu-

liert (1, 7), die Aufgabenformulierung und Absprachen, Struktur und Design, Datenerhebung/-aufbereitung, Mitteilung und Veröffentlichung sowie die Umsetzung der Evaluationsergebnisse beschreiben. Wichtige Postulate hieraus sind u. a. die Angemessenheit und Transparenz des Evaluationsvorgehens, die Unabhängigkeit der Evaluatoren und die Einbeziehung der Evaluierten.

Derartige Standards legen jedoch das Design der Evaluation bzw. der Begleitforschung nicht fest. So bleibt die methodische Frage offen, ob die Evaluationsergebnisse eher in einen internen Lern- und Adaptationsprozess (Qualitätskreislauf) münden sollen (damit allerdings von geringer externer Validität und Aussagekraft bleiben) oder ein strenges Forschungsdesign mit Kontrollgruppe, möglicherweise Randomisierung usw. gewählt werden soll, das allerdings nur außerordentlich schwierig zu realisieren ist.

In der Evaluationstheorie (11) gibt es eine Unterscheidung unterschiedlicher Evaluationstypen. Im Rahmen der Evaluation von DMPs sind dabei die Prozessevaluation (formative E.), die Wirkungsanalyse (summative E.) und die Programmkontrolle die wichtigsten Elemente. Eine ‚front-end analysis‘ – eigentlich der früheste Evaluationsschritt, in dem die Problemlage und die Effektivität der einzusetzenden Programmelemente analysiert wird – meinen die politisch Entscheidenden bereits hinter sich zu haben, wenn auch anzumerken ist, dass eine Bedarfsanalyse, feasibility studies und eine Bewertung einzelner Programmelemente bisher nicht mit der notwendigen Gründlichkeit durchgeführt worden ist. Aufsätze wie der vorliegende versuchen nun, wenigstens einen weiteren wesentlichen Schritt, das ‚evaluability assessment‘ nachzuholen. Demgegenüber liegt der sechste grundlegende Typ, die vergleichende Evaluation von Evaluationen (Metaevaluation) noch in beträchtlicher Ferne (Evaluationstypen nach 6).

Stufenmodell der Evaluation komplexer Interventionen

Einen Ausweg aus dem methodischen Dilemma unterschiedlicher Studiendesigns bietet ein Vorschlag des britischen Medical Research Council, nach dem komplexe Interventionen ein gestuftes evaluatives Vorgehen erfordern (4). Ähnlich wie in der Arzneimittelentwicklung – von der präklinischen Studie am Labormodell bis zur Phase-IV-Studie nach Markteinführung – in zunehmendem Maße die Komplexität des Anwendungsbezuges sowie praxisergebnisgerechte Outcomeparameter in das Studiendesign eingeschlossen werden, sollte auch die Evaluation komplexer gesundheitspolitischer Interventionen gestuft durchgeführt werden (vgl. Abb. 1).

Beispielhaft haben Mühlhauser und Berger (10) ein solches gestuftes Vorgehen anhand von Untersuchungen zur Patientenschulung bei Diabetes und Hypertonie dargestellt. Entscheidend ist dabei, dass grundsätzliche Erkenntnisse zur Wirksamkeit der Einzelkomponenten der Intervention bereits in frühen Phasen der Programmformulierung und in relativ einfachen Studiensituationen gewonnen werden. In späteren Evaluationsphasen (parallel zur Implementierung der DMPs) können entsprechend komplexere Studienfragestellungen und Designs gewählt werden.

Folgt man dem Vorschlag von Campbell et al. (4), müssen die Einzelkomponenten sowie die Anlage des Gesamtprogramms bereits in Phase I/II hinreichend untersucht worden sein, um in Phase III bereits eine – durchaus kontrolliert angelegte – Evaluationsstudie durchzuführen. Im Fall der DMPs wäre dies der Zeitpunkt der regional begrenzten Pilotvorhaben. Für die ersten beiden Programme (Typ 2-Diabetes, Brustkrebs) ist dieser Zeitpunkt bereits verstrichen. Für die derzeit in Arbeit befindlichen könnte dies noch erfolgen. Phase III-Studien wären vom Anspruch an die Zielpa-



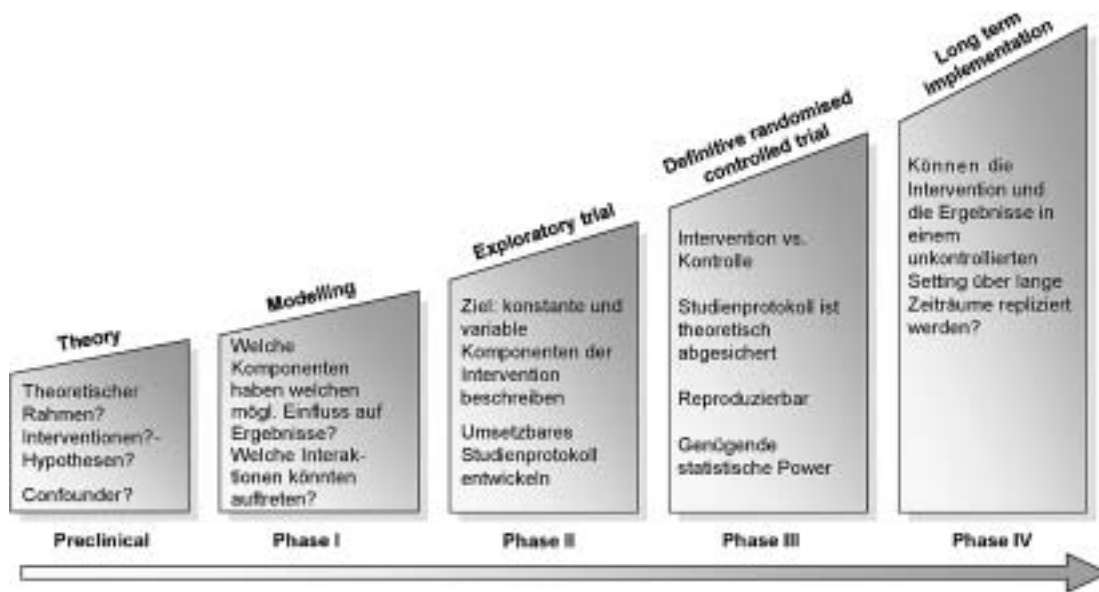


Abb 1. Design und Evaluation komplexer Interventionen (modifiziert nach Campbell M et al. BMJ 2000; 321: 694–696).

parameter und die Zeitperspektive nicht zu überfrachten, auch kann man hier ggf. gut belegte Surrogatparameter beobachten. Sicher ist heute schon, dass hier Vollerhebungen aus wissenschaftlicher Sicht unnötig sind. In ihrer Basis, Ziehung und Umfang genau geplante Stichproben reichen aus und sparen Kosten.

Der Fokus in Phase IV liegt auf dem Transfer in die Versorgungspraxis, der Implementierung der Programmbestandteile, Vergleichsbewertung unterschiedlicher Programme bzw. Leistungserbringer, sowie der Akzeptanz und längerfristigen Wirksamkeit strukturierter Versorgung. Während in dieser Phase nicht zwingend mit Kontrollgruppen gearbeitet werden muss, können auch sich erst längerfristig manifestierende ‚harte Outcomes‘ (wie die Niereninsuffizienz beim Diabetes) zum Evaluationskriterium werden. Der Beteiligung der Evaluierten, der Steuerung des Implementierungsprozesses und dem Zusammenspiel der Programmkomponenten, Anreizen etc. kann hier besondere Aufmerksamkeit geschenkt werden. Insbesondere in dieser Phase sollten Evaluationen mit einem zeitnahen Feedback an die Leistungserbringer verbunden werden.

Kurz gesagt, geht es bei diesen Studien während der Breitereinführung der DMPs nicht darum, zum wiederholten Male nachzuweisen, dass einzelne Elemente von strukturierter Versorgung, Patientenschulung etc. als solche wirksam sind, sondern dass die Implementierung eines gesundheitspolitischen Programms durch die verschiedenen Leistungserbringer angemessen funktioniert und messbare Wirkungen auf die gesundheitliche Situation der eingeschlossenen Patientengruppe hat (und natürlich die Frage, ob die Mehrzahl der potentiell betroffenen Patienten auch erreicht wird).

SCHLUSSFOLGERUNGEN

- Die Evaluation von Disease-Management-Programmen ist notwendig und hierzulande gesetzlich vorgeschrieben.
- Bisher besteht noch weitgehende Unklarheit über die konkrete Ausgestaltung geeigneter Strategien zur Programmevaluation.
- Anhand der wissenschaftlichen Begleitung der Diabetesvereinbarungen in Nordrhein werden exemplarisch aktuelle Defizite in der Evaluation bereits bestehender Programme für chronisch Kranke deutlich.

• Auf der Basis internationaler Erfahrungen (u.a. 8) ergeben sich in folgenden Dimensionen sinnvolle Anforderungen an eine Programmevaluation:

- Nützlichkeit: Ausrichtung an Informationsbedürfnissen der Nutzer
 - Durchführbarkeit: praktikabel, gut durchdacht, diplomatisch, zu vertretbaren Kosten
 - Korrektheit: rechtlich und ethisch korrekt, Offenlegung von Ergebnissen
 - Genauigkeit: valide, reliable Ergebnisse, begründete Schlussfolgerungen, formative und summative Metaevaluation
- Als Ausweg aus dem methodischen Dilemma unterschiedlich geeigneter Studiendesigns wird nach dem Vorbild des britischen Medical Research Council ein vierstufiges Modell vorgeschlagen, das eine angemessene Evaluation auch komplexer Interventionen erlaubt.

SUMMARY

Evaluating of Disease Management Programmes – Current Deficits, Demands and Methods

The evaluation of disease management programmes is necessary and has been



made obligatory in Germany by the Fifth Book of Social Code. So far there has been extensive ambiguity as regards the precise definition of suitable strategies for the evaluation. Considering the scientific evaluation of the diabetes contract in North-Rhine current deficits of the evaluation of already existing programmes for chronically ill patients are exemplarily described. On the basis of international experience reasonable demands are discussed that ought to be accomplished when evaluating the effectiveness and benefit of the programmes. Taking the British Medical Research Council as an example, a four-step model will be suggested as a loophole in the methodological dilemma of varying suitable study designs, which allows for the adequate evaluation of even complex interventions.

Key words: disease management, programme evaluation, evaluation methodology

LITERATUR

1. American Evaluation Association (2003) www.eval.org/Publications/publications.html, acc. 29.04.2003
2. Altenhofen L, Haß W, Oliveira J, Brenner G (2002) Modernes Diabetesmanagement in der ambulanten Versorgung. Ergebnisse der wissenschaftlichen Begleitung der Diabetesvereinbarungen in der Kassenärztlichen Vereinigung Nordrhein. Wissenschaftliche Reihe des Zentralinstituts für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland. Band 57 Köln, Deutscher Ärzte-Verlag
3. Bodenheimer T, Wagner EH, Grumbach K (2002) Improving primary care for patients with chronic illness. *JAMA* 288: 1775–1779, part 2: 288: 1909–1914
4. Campbell M, Fitzpatrick R, Haines A, et al. (2000) Framework for the design and evaluation of complex interventions to improve health. *BMJ* 321: 694–696, vgl. auch: www.mrc.ac.uk/prn/pdf-mrc_cpr.pdf, acc. 30.04.2003
5. Campbell N, Murchie P (2003) The international perspective of structured disease management programmes for secondary prevention in coronary heart disease. In: Medizinischer Dienst der Spitzenverbände der Krankenkassen e.V. (ed) Grundlagen einer evidenz-basierenden ambulanten Versorgung von KHK-Patienten. Essen (im Druck)
6. Chelimsky E (1990) Evaluation und Politik. In: Koch U, Wittmann WW (eds) Evaluationsforschung. Berlin: Springer, 249–280
7. Gerlach FM (2001) Qualitätsförderung in Praxis und Klinik. Stuttgart, Thieme, 239–248
8. Joint Committee on Standards for Educational Evaluation (o.J.) The program evaluation standards. www.wmich.edu/evalctr/jc/, acc. 30.04.2003
9. Mühlhauser I, Berger M (1996) Surrogat-Marker: Trugschlüsse. *Deutsches Ärzteblatt* 93: A3280–A3283
10. Mühlhauser I, Berger M (2002) Patient education – evaluation of a complex intervention. *Diabetologia* 45: 1723–1733
11. Øvretveit J (2002) Evaluation gesundheitsbezogener Interventionen. Bern, Huber
12. Rychetnik L, Frommer M (2002) A schema for evaluating evidence on public health interventions. Version 4. National Public Health Partnership, Melbourne, Australia. <http://hna.ffh.vic.gov.au/nphp/ppi/evidence/schema>, acc. 29.04.2003
13. Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (2003) Finanzierung, Nutzerorientierung und Qualität. Gutachten 2003. Onlineversion (Langfassung) www.svr-gesundheit.de/gutacht/sogu03/03deut/lang03.pdf, acc. 29.04.2003
14. Task Force on Community Preventive Services (2002) Recommendations for healthcare system and self-management education interventions to reduce morbidity and mortality from diabetes. *American Journal of Preventive Medicine* 22: 10–14
15. WHO (2001) International Classification of Functioning, Disability and Health, ICF www3.who.int/icf/icftemplate.cfm, acc. 30.04.2003

Korespondenzadresse:

Prof. Dr. med. Ferdinand M. Gerlach, MPH, Institut für Allgemeinmedizin, Universitätsklinikum Schleswig-Holstein, Campus Kiel, Arnold-Heller-Straße 8, D-24105 Kiel, Germany
Tel.: 0431-597-2226;
Fax: 0431-597-1183;
e-mail: gerlach@allgemeinmedizin.uni-kiel.de

